Electron Transport Using the Quantum Corrected Hydrodynamic Equations

J.P. KRESKOVSKY* and H.L. GRUBIN*

Scientific Research Associates, Inc., Glastonbury, CT 06033

Transport in one- and two-dimensional semiconductor device structures is considered using a set of quantum corrected hydrodynamic equations. Simple one-dimensional simulations demonstrate the need to include quantum effects in structures with sharp interfaces. Application to a two-dimensional quantum well HEMT structure is then considered. A brief discussion of the computational procedure is also presented.

Key Words: *Quantum-Hydrodynamics, HEMT, InAlAs / InGaAs*

INTRODUCTION

The dimensions of semiconductor devices have continued to shrink as a result of advances in crystal growth and processing technique. With these advances and reductions in size new families of devices employing sharp interfaces have emerged. These factors have created the need to routinely include quantum mechanical effects in the analysis of device operation. Such quantum effects appear in structures when significant variations in density occur over length scales of the order of the thermal de Broglie wavelength. Situations such as these arise in many devices which are common today. These include devices where transport is confined to narrow channels through heterojunctions and quantum wells, such as HEMTs, and devices where transport is across barriers, such as resonant tunneling diodes (RTDs). In fact, one can safely say that in any region of a device where a high concentration of carriers is present near an "abrupt" heterojunction, quantum mechanical effects will be important. Fortunately, predictive capabilities have also continued to grow in sophistication and the ability to simulate quantum mechanical phenomena in these new structures is now becoming a reality. This is not to say

that the modeling problems have been solved, but rather that significant progress has and is being made. Quantum effects have been simulated using the single particle Schrodinger's equation, the Liouville equation or the Wigner equation. While these equations are rigorous, application to full two-dimensional device structure can be unruly (from a simulation viewpoint). Thus, there remains a need for a transition from the classical to the quantum regime which is filled through the use of a set of quantum corrected hydrodynamics equations (QHD). These equations stem from the classical hydrodynamic equations with additional terms arising from corrections including the Bohm [1] and Wigner [2] potentials.

Various developments of the quantum hydrodynamic equations have appeared in the literature [3, 4, 5]. The form we employ in our simulation procedure is given as, in one dimension,

$$\frac{\partial n}{\partial t} = -\frac{\partial nu}{\partial x} \qquad (1)$$

$$\frac{\partial u}{\partial t} = -\frac{\partial u^2/2}{\partial x} - \frac{k_0}{m}\frac{\partial T}{\partial x}$$

$$+ \frac{e}{mT}\left(\psi + \frac{Q}{3} + \psi_c - \psi_f\right)\frac{\partial T}{\partial x} + \frac{e}{m}\frac{\partial \psi_f}{\partial x} - \frac{u}{\tau_m} \qquad (2)$$

$$\frac{\partial}{\partial t}\left(\frac{3}{2}k_0 T - \frac{Q_e}{2}\right)$$

$$= -u\frac{\partial}{\partial x}\left(\frac{3}{2}k_0 T - \frac{Q_e}{2}\right) - (k_0 T - Q_e)\frac{\partial u}{\partial x}$$

$$+ \frac{1}{n}\frac{\partial}{\partial x}\left(\kappa\frac{\partial T}{\partial x}\right) - \frac{(T - T_0)}{\tau_e} + \frac{u^2}{\tau_m} \qquad (3)$$

and

$$\frac{\partial}{\partial x}\left(\varepsilon\frac{\partial \psi}{\partial x}\right) = e(n - N_D). \qquad (4)$$

The Bohm potential is defined as

$$Q = \frac{h^2}{2m}\frac{1}{\sqrt{n}}\frac{\partial^2\sqrt{n}}{\partial x^2} \qquad (5)$$

and the Wigner potential is given as

$$Q_e = \frac{h^2}{12m}\frac{\partial}{\partial x}\left(\frac{1}{n}\frac{\partial n}{\partial x}\right). \qquad (6)$$

An "equation of state" is introduced as

$$n = N_c \exp\left(-(E_t - E_f)/k_0 T\right)$$

$$= N_c \exp\left(\psi + \psi_c + \frac{Q}{3} - \psi_f\right)\frac{e}{k_0 T} \qquad (7)$$

for Boltzmann statistics. Here, n is the electron density, u is the electron velocity, T is the temperature of the electron gas, m is the effective electron mass, N_D is the donor doping level, N_c is the electron density of states, ψ is the self-consistent electrostatic potential, and τ_m and τ_e are relaxation times introduced to model scattering. The potential, ψ_c, is introduced to account for the variations in the band structure of different materials while ψ_f is related to the Fermi level as $e\psi_f = -E_f$. The energy E_t in equation (7) is given as

$$E_t = E_c - \frac{eQ}{3} = -e\left(\psi + \frac{Q}{3} + \psi_c\right) \qquad (8)$$

and is the effective conduction band energy level including corrections for the Bohm quantum potential.

As will be discussed subsequently, the form of equations (1–7) is, in part, based on consideration of the numerical solution procedure and in part for physical reasons. Equation (3) is developed from the more familiar form of the energy equation by taking the moment of the momentum equation with respect to velocity, obtaining a mechanical energy equation, and subtracting it from the total energy equation. The present form of Eq. (1) is obtained by substituting Eq. (7) for n in the pressure gradient term (also referred to as the diffusion term), expanding, and eliminating the field terms which include gradients of the conduction band energy and quantum potential. This form is desirable since, in the absence of an applied bias the Fermi level, ψ_f, will be constant and we recover the equilibrium condition $T = T_0$, $u = 0$ with great accuracy in our simulations. Additionally, when a bias is applied the Fermi-level will remain continuous and the gradient of the Fermi-level appearing in Eq. (2) may be differenced in a straightforward manner. The discontinuities associated with variations in the conduction band energy level appear only in a coefficient of the temperature gradient. Thus, we may treat heterojunctions as abrupt interfaces if desired without a loss of accuracy or special differencing techniques.

The collision terms have been treated classically through the introduction of relaxation times. The relaxation times would ideally be functions of momentum and energy. However, at this stage of the development of our simulation procedure, we have related the relaxation times to the uniform field velocity-field curve. Under the condition of a uniform field in the momentum equation, Eq. (2), reduces to

$$u = \frac{\tau_m e}{m}\frac{\partial \psi_f}{\partial t} \equiv \mu\frac{\partial \psi_f}{\partial x} \qquad (9)$$

where we define μ as the mobility of the material in question. From Eq. (9) we arrive at the relationship

$$\tau_m = \frac{m\mu}{e} \qquad (10)$$

To account for the energy dependence of the relaxation time we note that under uniform field conditions a specific value of field leads to a unique value of temperature. Thus, we relate the energy dependence of the relaxation time to a field dependence of the mobility. This field dependence is given in the present simulations as

$$\mu(E) = \mu_0 \left/ \sqrt{1 + \left(\frac{\mu_0(|E| - |E_0|)}{V_s}\right)^2}\right. \qquad (11)$$

where μ_0 is the low field mobility, V_s is the satu-

rated drift velocity, E is the electric field and E_0 is the zero bias or equilibrium field. The equilibrium field must be subtracted from the self-consistent field since it does not contribute to an increase in carrier energy. The energy relaxation time is related to the momentum relaxation time.

This approach may be satisfactory for structures such as HEMTs where quantum effects serve to influence the charge sheet density and distribution while transport is parallel to the interfaces, thus classical in nature. However, for structures such as RTDs, the treatment is incomplete and is presently the source of considerable study.

It is also important to have a sound idea of the length scales over which quantum effects are important. This is necessary in determining a suitable grid structure in a simulation, as well as to provide insight with regard to the physics of a given problem. We recognize from both Eqs. (2 & 7) that if the quantum potential is to have a first order effect on momentum transport or density at an interface, then it must be of the same order of the potential difference in the conduction band, $\Delta \psi_c$. From Eq. (5) we see that this requires that $\hbar^2/(2mx^2\Delta\psi_c)$ must be of order one. Using typical values for m and $\Delta\psi_c$, a length scale of the order of 10 Å emerges. Thus, if quantum effects are to be important, interfaces must be very sharp. Grading interfaces over 100 Å or more will significantly reduce the transport due to quantum effects. This analysis also tells us that if we are to accurately resolve quantum effects in our simulations, we must employ computational grids with spacings of the order of one angstrom at abrupt interfaces. In the simulations presented here we have used grid spacings as small as 0.5 Å in our 1-D computations and 4.0 Å in the 2-D results, to ensure accurate resolution.

THE COMPUTATIONAL PROCEDURE

Equations (1)–(7) form a coupled system of nonlinear partial differential equations which must be solved. In order to solve these equations numerically the equations must be linearized and then the resultant system must be discretized in space and time. We note that as written the highest order derivative which appears is of second order. Thus, three point spatial differencing approximations can be applied and, in one dimension, the resulting matrix system of linear equations would be block tridiagonal in structure. There exist straightforward procedures for solving block, tridiagonal systems and, in this sys-

tem, boundary conditions are easily applied. We note also that the Bohm and Wigner potential could be substituted directly into the remaining equations with a reduction in the number of equations to be solved from seven to five. However, the order of the highest derivative now is raised to third and derivatives will appear in the exponent in Eq. (7). The higher order derivative will increase the bandwidth of the matrix system to pentadiagonal if a centered difference approximation is used and spurious boundary conditions will be required for the electron density. While these boundary conditions can be developed from specification of Q and Q_e on the boundaries, the added bandwidth can become cumbersome. Thus, we chose to retain the equations for Q and Q_e and solve the coupled tridiagonal system of seven equations.

To solve the system of Eqs. (1)–(7) in one dimension we apply a linearized block implicit (LBI) procedure [6]. We have successfully used this procedure in single and multidimensional simulations using the drift and diffusion equations and moments of the Boltzmann transport equation [7, 8]. The system of equations (1)–(7) may be expressed as

$$\frac{\partial H(\phi)}{\partial t} = D(\phi) + S(\phi) \qquad (12)$$

where

$$\phi = \left[n, u, T, \psi, Q, Q_e, \psi_f \right]^T \qquad (13)$$

and $H(\phi)$, $D(\phi)$, and $S(\phi)$ represent functions of the dependent variables. $D(\phi)$ represents the spatial operators and $S(\phi)$ source terms. The equations are linearized, for example, as

$$G(\phi)^{n+1} = G(\phi)^n + \Delta t \frac{\partial G(\phi)}{\partial \phi} \left. \frac{\partial \phi}{\partial t} \right|^{n+1} + O(\Delta t^2)$$

$$\approx G(\phi)^n + \left. \frac{\partial G(\phi)}{\partial \phi} \right|^n \Delta\phi^{n+1} + O(\Delta t^2)$$

$$(14)$$

where the super script indicates a time level,

$$t^{n+1} = t^n + \Delta t. \qquad (15)$$

Performing the linearizations and collecting terms results in the equation

$$(A + \Delta t L)\Delta\phi^{n+1} = \Delta t [D(\phi^n) + S(\phi^n)] \qquad (16)$$

where

$$A \equiv \left(\frac{\partial H}{\partial \phi} - \Delta t \frac{\partial S}{\partial \phi} \right)^n \qquad (17)$$

and

$$L \equiv -\frac{\partial D}{\partial \phi} \bigg|^n . \qquad (18)$$

Equation (16) represents a 7 × 7 block coupled system of linear partial differential equations. Introducing a computational mesh of N grid points and employing three point centered difference operators to approximate the operator L, equation (16) may be expressed as

$$M\Delta\phi = B \qquad (19)$$

where M is an $N \times N$ block tridiagonal matrix with 7 × 7 block size. The narrow banded block structure allows for efficient solution of Eq. (19) using direct methods.

When multidimensional problems are considered the narrow band structure of the M matrix is lost. If a two-dimensional problem is considered on an $N \times N$ mesh then the M matrix will be of rank N^2 and the bandwidth will be N. However, the matrix may be split using ADI techniques [9] to reduce the two dimensional matrix to a series of one-dimensional matrices. Letting the L operator be expressed as the sum of the x and y direction operators,

$$L = L_x + L_y \qquad (20)$$

equation (16) can be split as

$$(A + \Delta t L_x)\Delta\phi^* = \Delta t [D(\phi) + S(\phi)]^n \qquad (21a)$$

$$(A + \Delta t L_y)\Delta\phi^{**} = A\Delta\phi^* \qquad (21b)$$

and

$$\Delta\phi^{**} = \Delta\phi^{n+1} + O(\Delta t^2). $$

Each of equations (21) is block tridiagonal and of rank N. Each is repeated N times in the x and y directions, respectively. The result is a significant reduction in computation at the expense of introduction of some additional error due to the splitting. As the solution converges, the splitting error goes to zero and the solution to Eq. (16) is recovered.

RESULTS

Simulations based on Eqs. (1)–(7) have been performed for a variety of structures [3, 5] including heterojunction barriers and RTDs. More recently we have begun simulation of a quantum well HEMT structure. Zhou and Ferry [10, 11] have also reported results on the effects of quantum transport in MESFETs using a similar system of quantum corrected hydrodynamic equations. Some of our results are compared to solutions based on the quantum Liouville equation here and in [5].

One-Dimensional Results

The first structure to be considered here is that of a simple 200 meV barrier structure in the center of a uniformly doped ($1 \times 10^{16}/cm^3$) 6000 Å long device. For this structure only equilibrium solutions were considered. These solutions are compared to those obtained from the solution of the quantum Liouville equation in Fig. 1 (from [5] with permission). Figures 1a and 1b show the density and potential energy when zero bias is applied. The field at the device boundaries is zero, thus the net space charge in the device is also zero. The variation of the density shows accumulation on the narrow gap side of the barrier equal to the depletion on the wide gap side (right). The band bending across the interface is shown in Fig. 1b. Both the density and the band structure agree well with the result from the Liouville equation. In Figs. 1c and 1d similar results are shown with a bias of -0.2 volts applied to the wide gap side of the junction. Again the agreement between the QHD equation and the Liouville equation is excellent.

In the previous simulation the quantum potential, Q, was divided by 3.0 in Eqs. 2 and 7. However, there is some discussion on what the actual value of the constant should be. To investigate the sensitivity of the QHD solutions to this factor a simulation was performed for a 2000 Å structure with a 300 meV barrier in the center. The narrow gap material was doped to $1 \times 10^{16}/cm^3$ while the wide gap side was doped to $1 \times 10^{18}/cm^3$. Results are shown in Fig. 2 (from [5] with permission) for values of the constant equal to 1.0, 3.0 and 9.0. We note that as the constant goes to infinity the classical result is obtained. In Fig. 2 we observe that as the value of the constant increases the peak density in the narrow gap side of the barrier also increases. From examination of the depletion region and the requirement
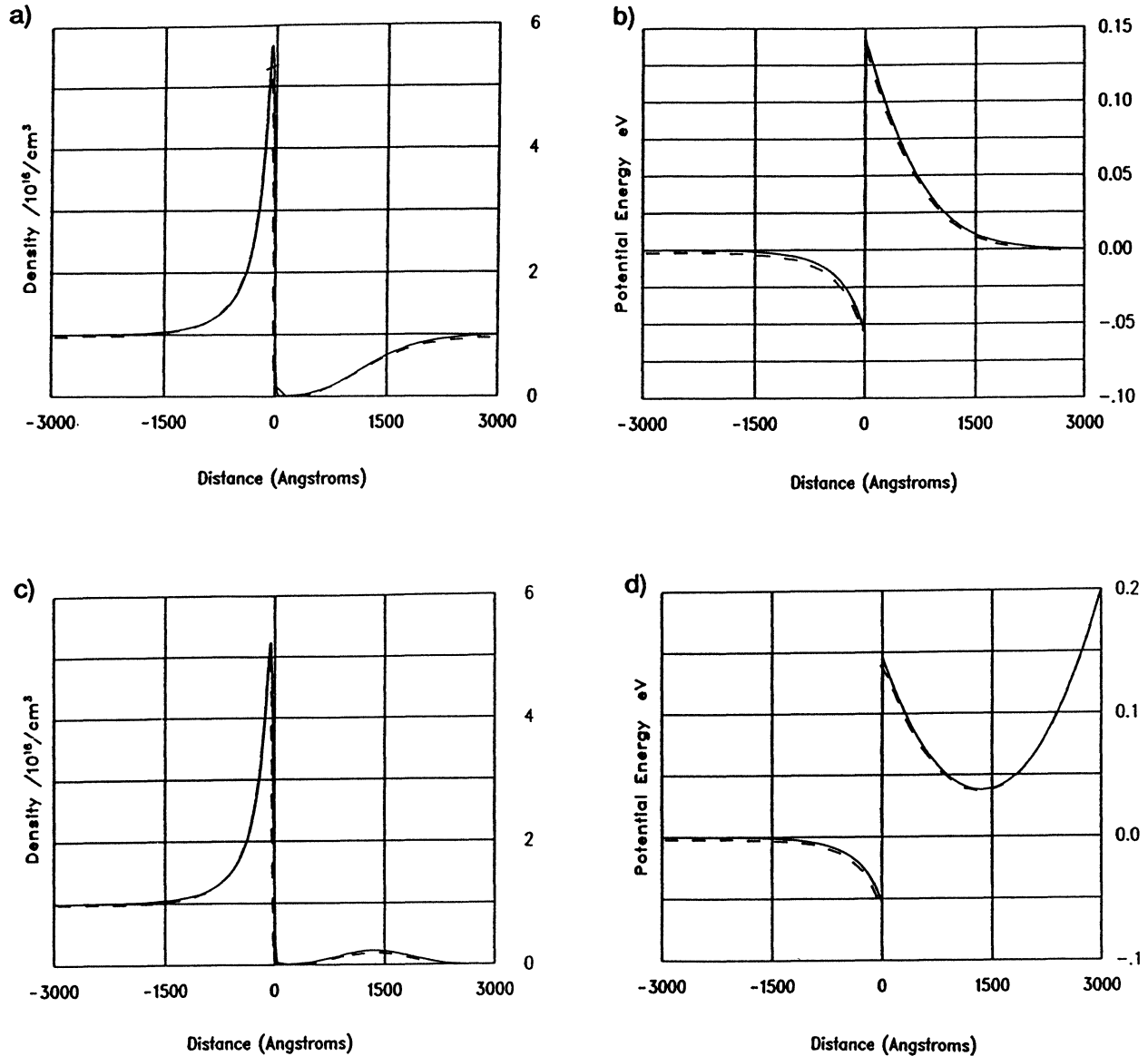
FIGURE 1 Results for a single 200 meV barrier structure. Solid line represents solution from the Liouville equation, dashed line from QHD equations. a) density, zero bias, b) potential energy, zero bias, c) density, −0.2 V bias, d) potential energy, −0.2 bias.

that the accumulation equal the depletion it is also evident that the higher the value of the constant, the greater the sheet charge density will be on the narrow gap side. Thus, this result indicates that a classical solution will over estimate the sheet charge density on the narrow gap side of structures with heterojunctions; e.g., HFETs, HEMTs, etc.

The final equilibrium simulations performed are for a one-dimensional slice taken through the region between the gate and source of a quantum well HEMT structure shown in Fig. 3. These results are shown in Figs. 4–7. Here solutions are compared with classical results (obtained by using $Q/1000$ in

Eqs. (2) and (7)). In the simulations the device is doped at a level of $2 \times 10^{18}/cm^3$ from 0 to 300 Å and nominally undoped from 300 Å to 10,000 Å. The actual structure is an InGaAs/InP HEMT with a 100 Å InGaAs cap layer followed by a 200 Å AlInAs gate isolation layer, a 30 Å AlInAs spacer layer, an 800 Å InGaAs channel, and a 2500 Å AlInAs buffer layer on an InP substrate (not shown). In the actual device the 200 Å gate isolation layer would be undoped with a planar doped layer at $5 \times 10^{12}/cm^2$ placed near the channel side of the isolation layer as shown in Fig. 3. However, the uniform doping is introduced for the preliminary 2D
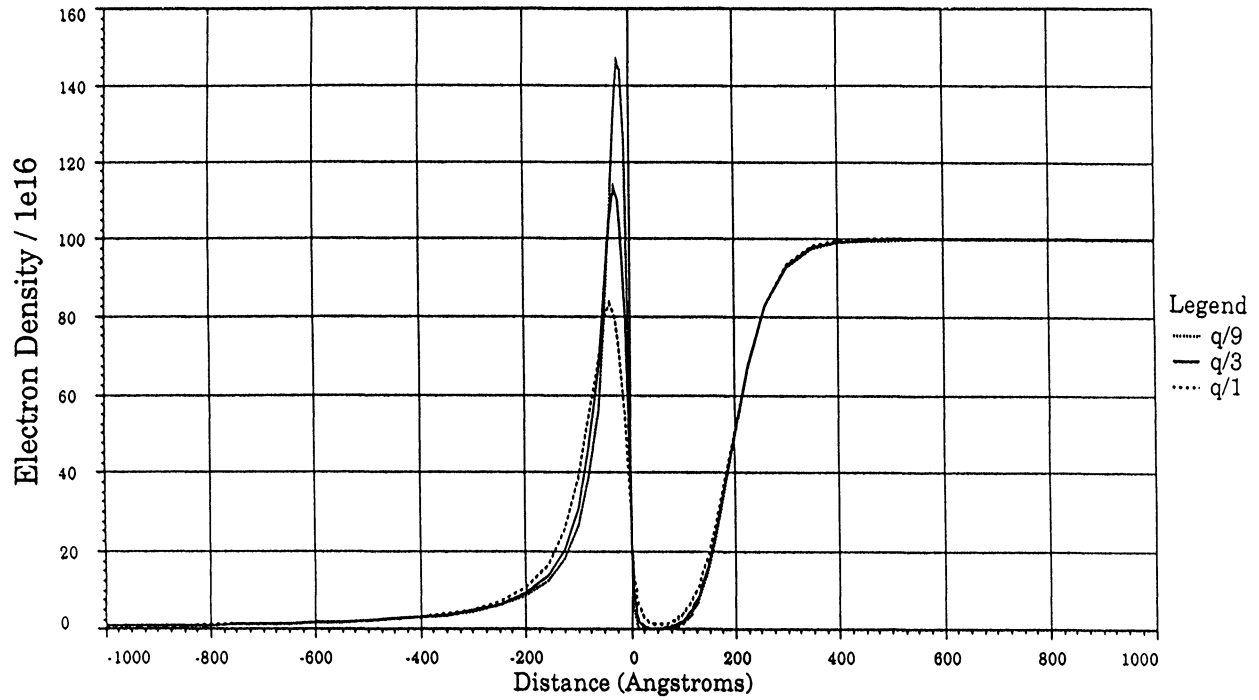
FIGURE 2   Variation of density for a 300 meV barrier structure as a function of the quantum correction constant.
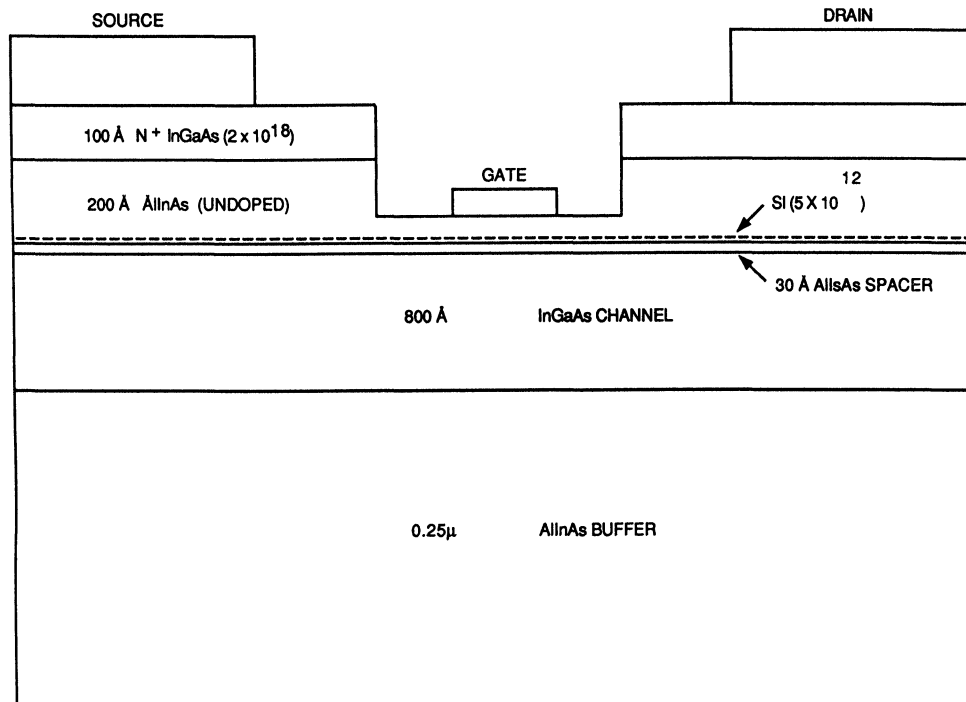

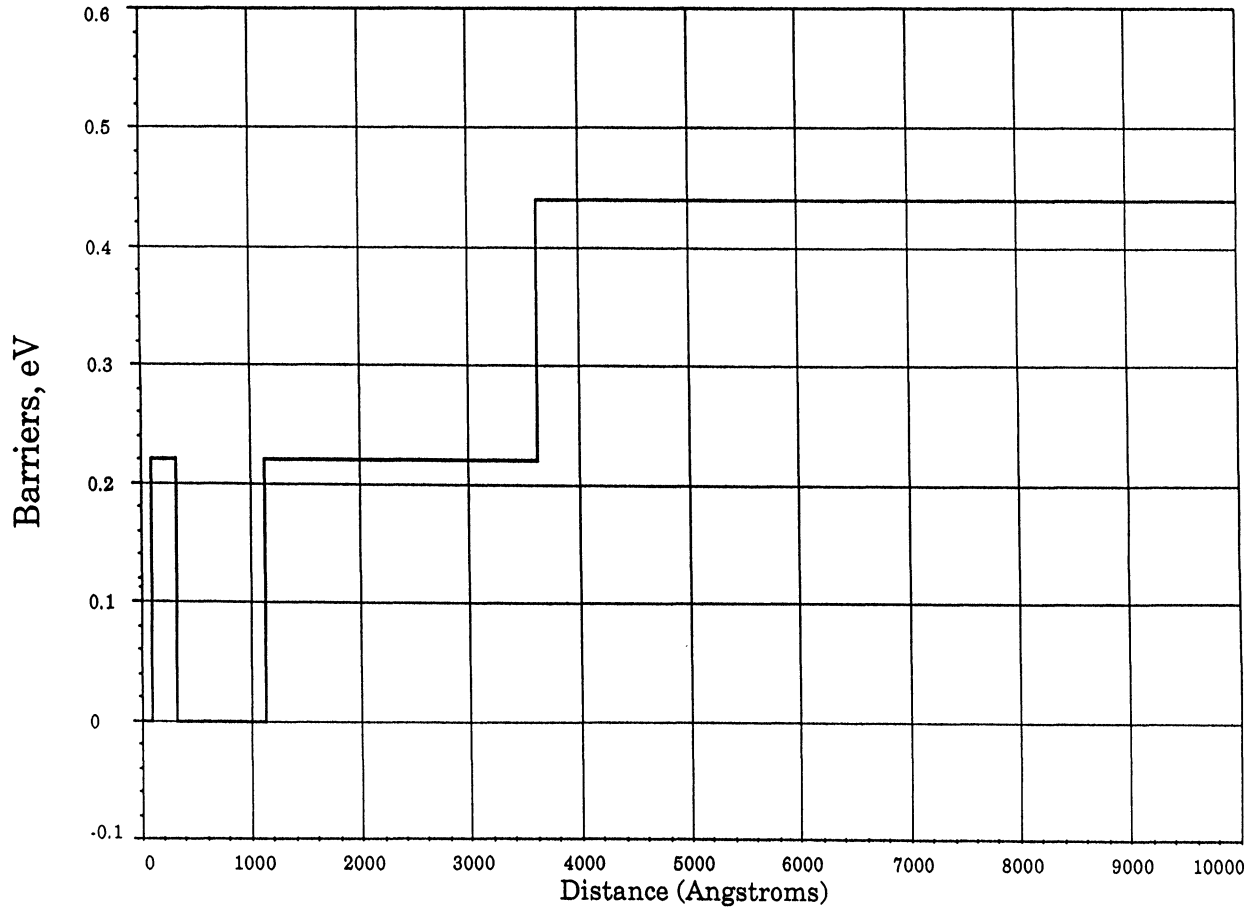FIGURE 3   Quantum well HEMT structure used in one and two dimensional simulations.

FIGURE 4   Barrier structure for direction normal to source contact for a quantum well HEMT.

simulations. We note that "the total charge" of the 200 Å layer doped to $2 \times 10^{18}/\mathrm{cm}$ is effectively $4 \times 10^{12}/\mathrm{cm}^2$. Thus the total charge in this layer is similar to that of the actual planar doped layer. The barrier structure across the device is shown in Fig. 4.

Figure 5 shows the electron density on linear (5a) and log (5b) scales over the first 2000 Å of the structure and compares the quantum mechanical result to the classical result. We see greater depletion of charge in the gate isolation layer in the classical result and an overprediction of the sheet density in the InGaAs channel. We also note the desired confinement of charge in the channel. The conduction band energy for the quantum and classical calculations are compared in Fig. 6a and the quantum mechanical result is compared to the conduction band total energy, which includes the quantum potential contribution, in Fig. 6b. Here we note from Fig. 6a that based only on $E_c - E_f$ the quantum result should show higher peak densities on each side of the gate isolation layer, but from Fig. 6b it is observed that the quantum potential

contribution lowers the barriers and yields a continuous variation of the conduction band energy across the interfaces. The quantum potential is shown in Fig. 7. Recall that 1/3 of this energy is added to the E_c to obtain E_t.

Two-Dimensional Results

Two-dimensional results have also been obtained for the InGaAs/InP HEMT structure shown in Fig. 3. The gate is recessed 100 Å into the gate isolation layer and the gate length is 1000 Å. Our initial simulations using a planar doped layer proved difficult in the two-dimensional case. These problems are related to convergence of the two-dimensional numerical procedure. Similar structures in one dimension posed no problems. To circumvent these convergence problems the structure was initially modified as in the one-dimensional case. The AlInAs gate isolation layer was doped. An additional modification to the structure was also intro-
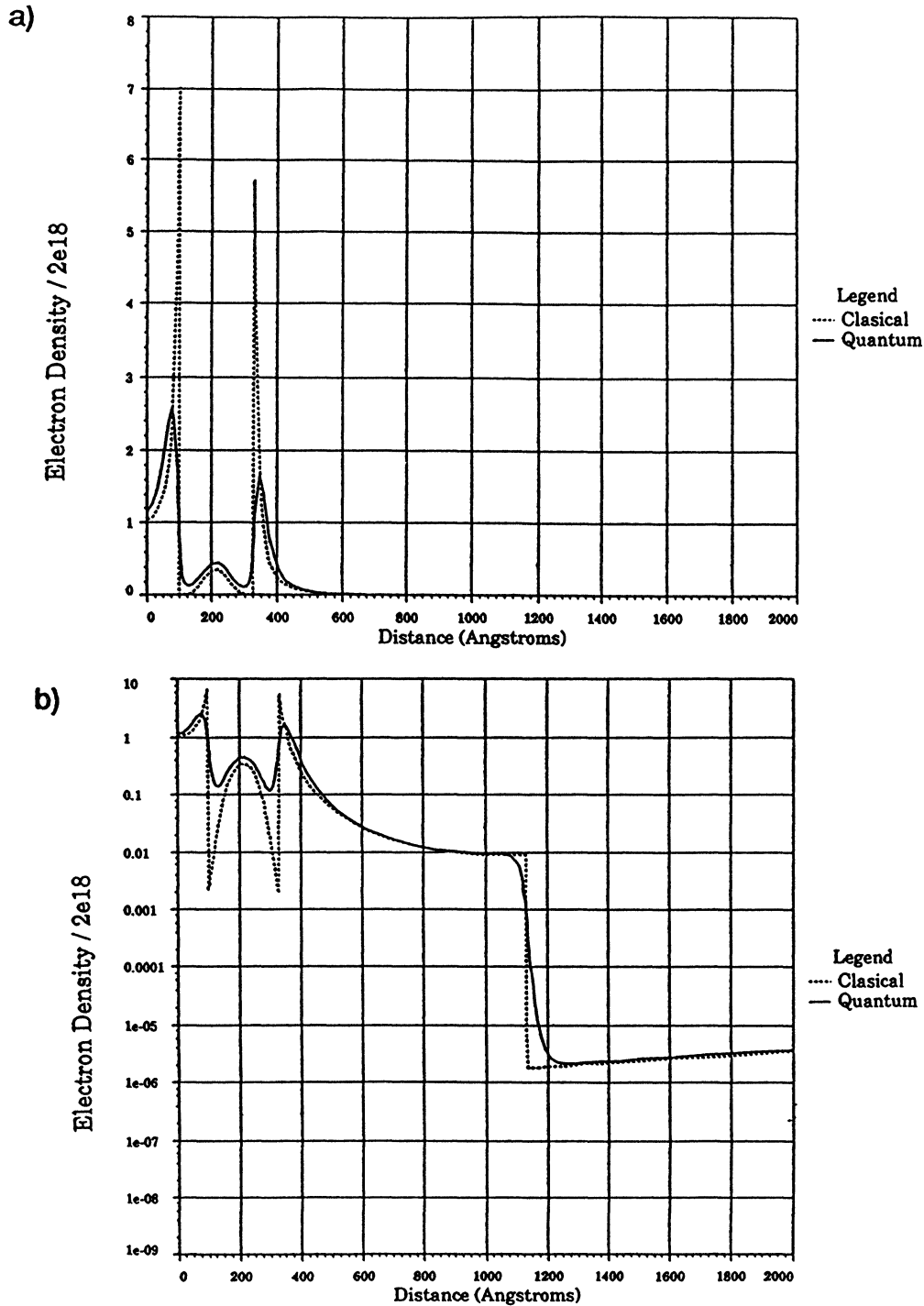
a)

b)

FIGURE 5   Comparison of QHD and classical solutions for the electron density normal to source contact surface in a quantum well HEMT, a) linear scale, b) log scale.

duced. In [12] it was observed experimentally that the I-V characteristics of a related device were similar, independent of whether or not the cap layer was doped. Some softening of the I-V characteristics was noted when the cap layer was doped but qualitatively the results were similar. It was noted in [12]

that through penetration of the metalization of the source and drain direct contact to the 2-DEG in the channel was achieved. The contacts were, as a result, ohmic with a reported resistance of approximately 0.3 $\Omega \cdot$ mm regardless of the doping of the cap layer. Thus, it would appear that the interfaces
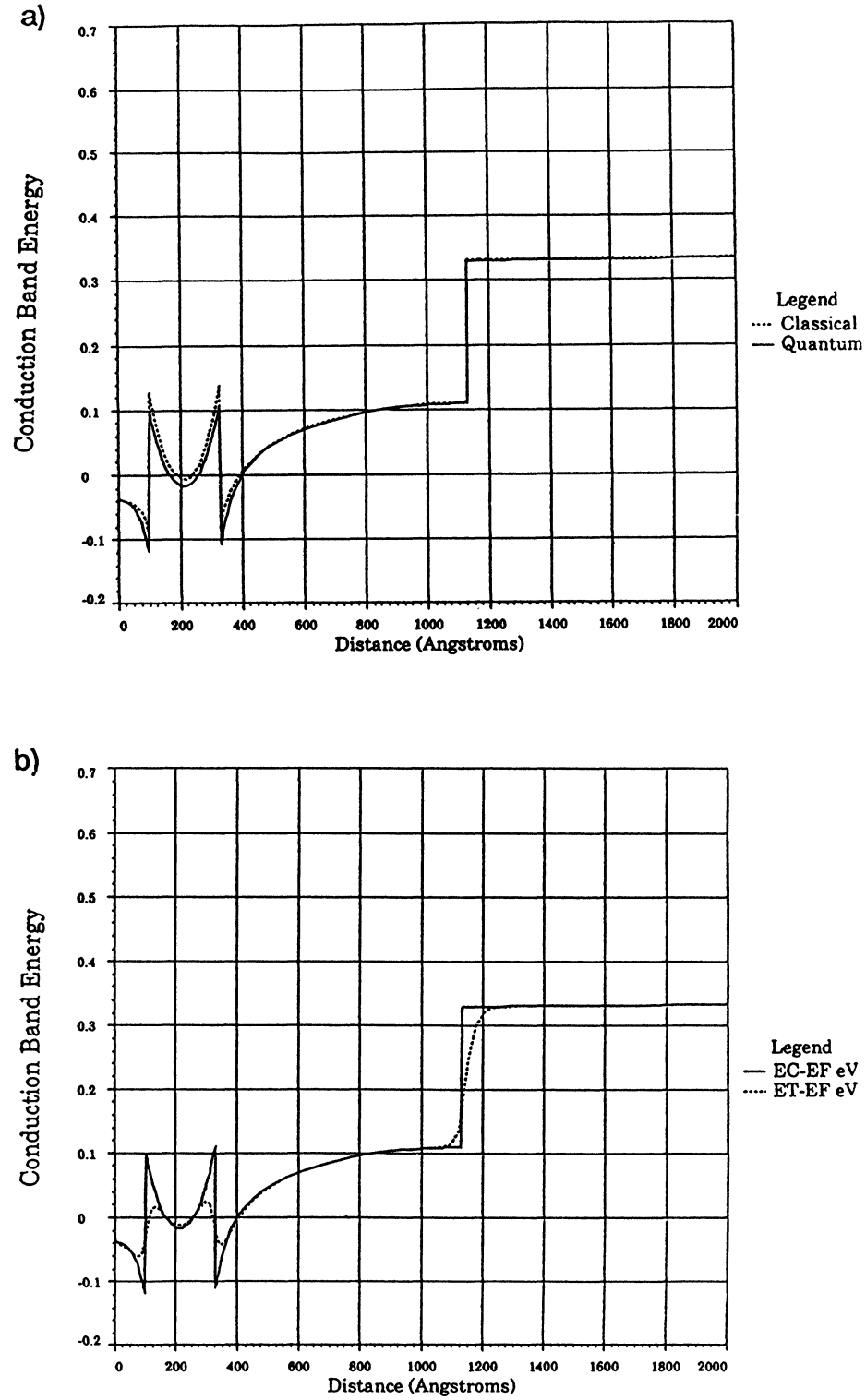
FIGURE 6   Conduction based energy distribution in a quantum well HEMT structure. a) comparison of QHD and classical result, b) comparison of QHD results for EC and ET, $ET = (EC + Q/3)$.
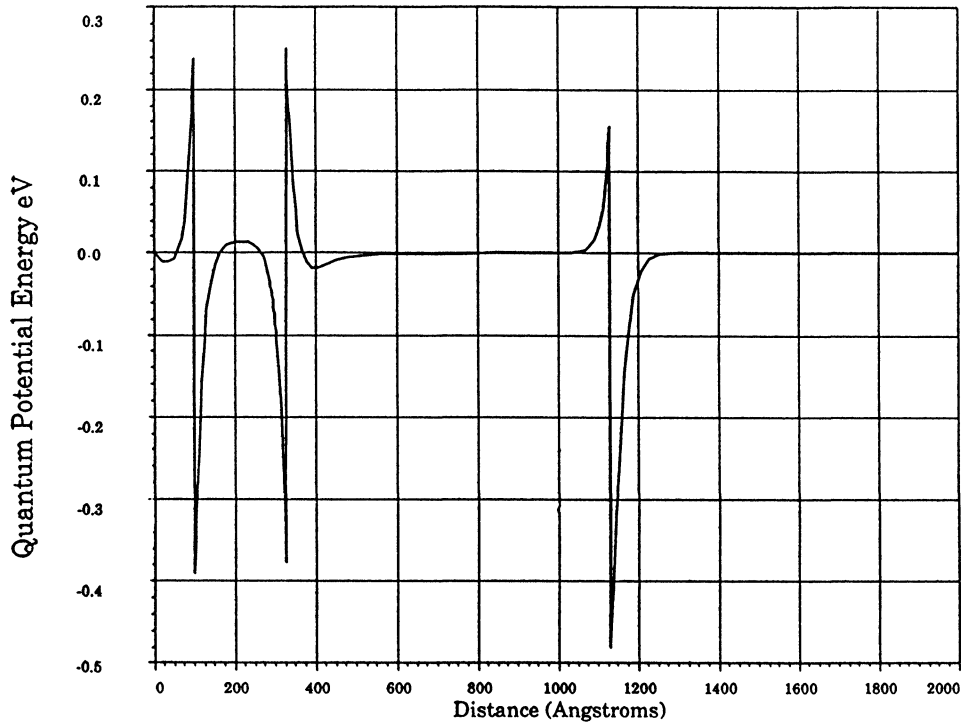
FIGURE 7   Variation of Quantum potential for the HEMT structure of Figure 3.

between the InGaAs cap layer, the AlInAs isolation layer and the channel were obliterated. Electrons could easily enter and exit the channel under the contacts as a result. However, away from the source and drain metalizations the heterojunctions would remain intact providing the desired isolation. In our simulations we have modeled this by elimination of the variations in the band gap between the various layers in the region of the source and drain and by extending the N^+ doping levels into the channel to mimic penetration by the metal. Without this modification, even with quantum corrections included in the governing equations, it was not possible to inject significant charge into the 2-DEG due to the barrier created by the gate isolation layer.

Typical results for this structure are shown in Figs. 8–11 for a bias of $V_{DS} = 0.75$ and $V_{gs} = 0.4$ volts. In Fig. 8 surface plots of the electron density, (a), potential (b), and electron temperature (c), are shown across the first 2000 Å of the device depth. The outline of the device structure is shown below the temperature surface for reference to later figures in which profile plots normal to the contact surface are presented. While the details of the distribution of electrons and potential are difficult to see in Fig. 8, the results do show interesting qualitative effects. We observe from the density surface the uniform contours under the source and drain result-

ing from the modeling of the metalization of the contacts, as discussed earlier. The potential surface also reflects this and the result is a low resistance ohmic contact to the 2-DEG in the channel. As we move away from the source or drain, towards the gate, an increase in electron concentration is observed with a sharp peak in the N^+ cap layer. Also, as a result of the applied drain bias we note the presence of a small field in the source-drain direction between the edge of the gate recess and the drain. This results in a lower electron density near the drain side of the gate recess as compared to that on the source side of the recess. While this will have little influence on the D.C. characteristics of the device, due to the function of the gate isolation layer, it will affect transient behavior. This "sloshing" of charge with bias level will affect the parasitic capacitance of the device. Near the source side of the gate recess we observe the presence of a second peak in the electron density. This peak is lower and is associated with the peak in the charge density in the 2-DEG in the channel. The peak runs the entire length of the channel, except at the source and drain regions, but cannot be clearly seen in this plot. The depletion of the 2-DEG under the gate is clearly evident in the results.

The electron temperature shows a sharp peak near the drain end of the gate. Over most of the
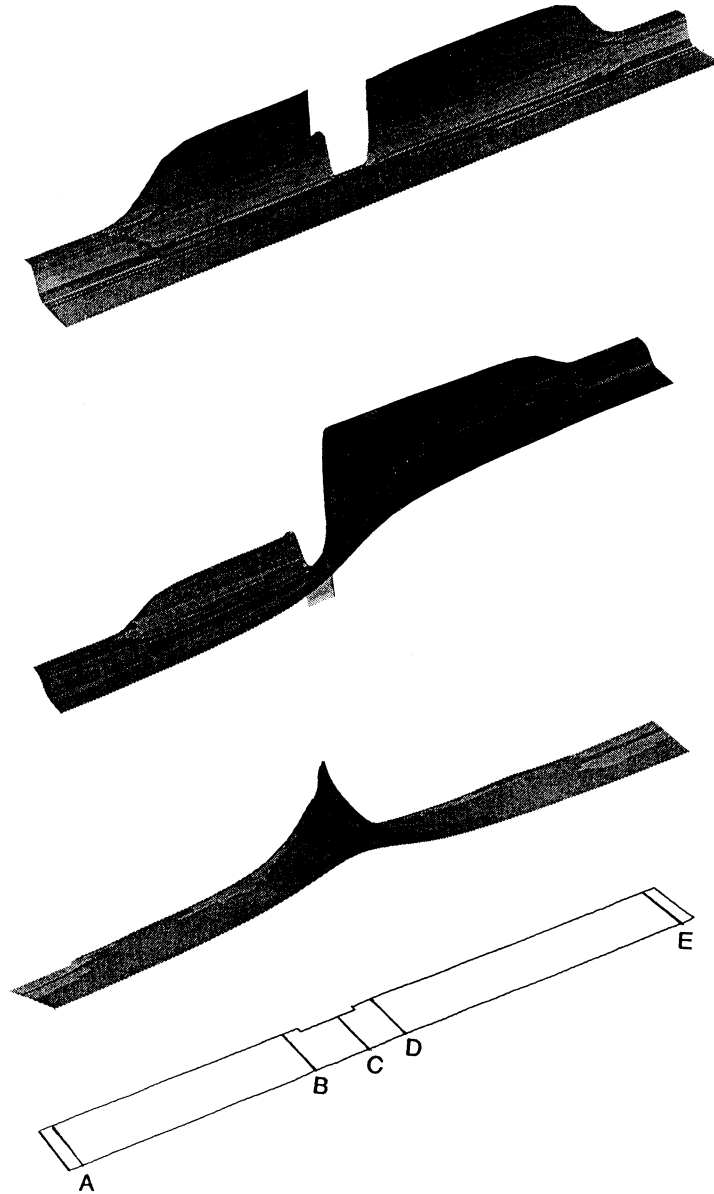
FIGURE 8   Surface plots of a) electron density, b) potential and c) electron temperature for a quantum well HEMT.

device the electrons are near equilibrium with the lattice. However, as the electrons accelerate in the low density region of the 2 DEG under the gate the temperature rises. The maximum temperature reaches approximately 600°K and occurs in the high field region just to the drain side of the gate. In this region the electrons slow down as they enter a high density region of the 2-DEG gas. The kinetic energy of the electrons is dissipated as heat in this region. As the electrons continue to decelerate, the temperature relaxes back to nearly the lattice temperature.

Due to the difficulty in deciphering the details of the solution from Fig. 8, we present profile plots of the same results in Figs. 9–11 along the lines marked on the projection of the device shown in Fig. 8. Figure 9 presents profiles of the electron density on both a linear (a) and log (b) scale. The curves labeled A & E in Fig. 9a show the electron density under the source and drain contacts. We note the penetration of the doping into the channel allowing direct contact to the 2-DEG. The curve labeled B shows the profile on the source side of the gate recess while curve D is on the drain side. Here we observe that the electron density in the channel has a similar peak level but electrons are drawn away from the substrate on the drain side due to the applied bias. We also note the higher peak density in the cap layer is on the source side of the gate
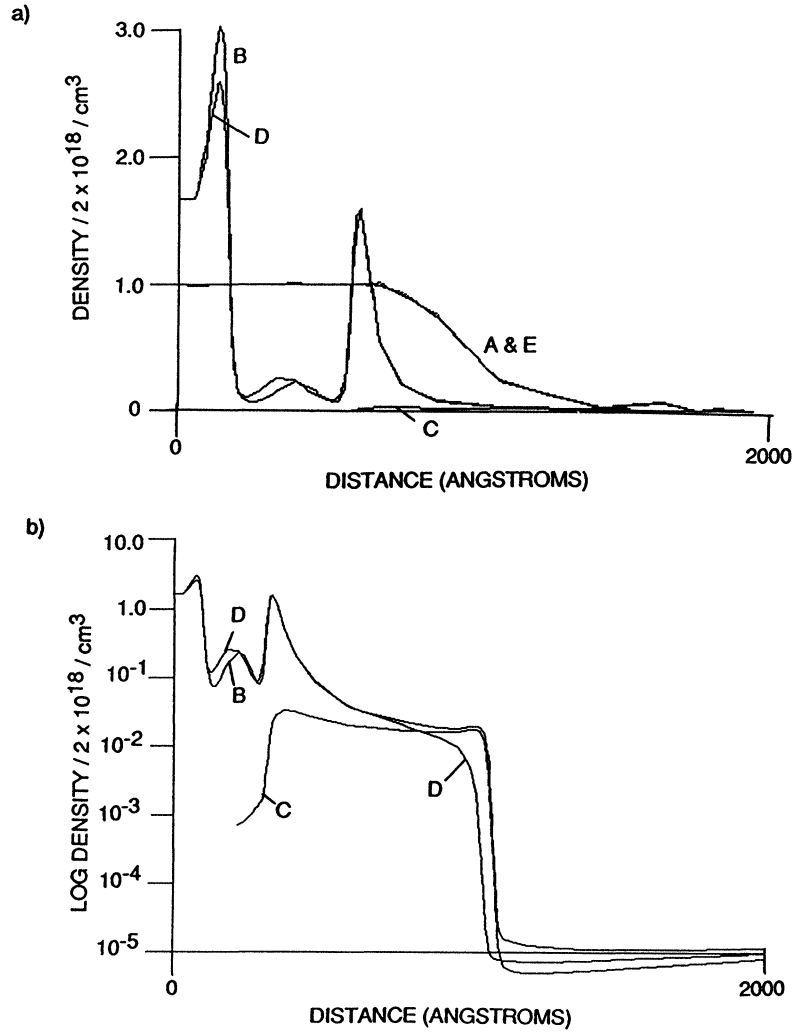
FIGURE 9   Electron density at selected planes normal to the device surface. a) linear scale, b) log scale.

recess, as discussed previously, due to the small source-drain direction field in this layer. Curve C represents the profile under the gate. We note here the significant depletion of the 2-DEG under the gate, even with a forward bias of 0.4 volts. We also note that the profiles are in good agreement with those obtained from our one-dimensional equilibrium solutions shown in Fig. 5, indicating that transport normal to the barrier structure is not significant and good isolation is achieved. Figure 10 shows the potential profiles at the same locations. The constant potential regions under the source and drain (A & E) again indicate low resistance contact to the 2-DEG in the channel and the peaks in potential curves (B & D) are a result of depletion of the charge in the isolation layer. Finally, profiles of electron temperature are shown in Fig. 11. Here we observe that under the source and drain the elec-

trons are close to being in equilibrium with the lattice (A & E). The maximum heating occurs on the drain side of gate (C), and the temperature then begins to relax back to the lattice temperature (D).

Having gained experience with our two-dimensional algorithm through simulation of the modified HEMT structure we returned to the original structure with an undoped 200 Å layer of AlInAs, a planar Si doped layer (assumed to be 10 Å thick and doped to $5 \times 10^{19}/cm^3$), and a 30 Å spacer layer between the cap layer and the channel. The region under the source and drain contacts was treated as in the modified structure except that the doping under the contacts was raised to $1 \times 10^{19}/cm^3$. Our initial result was to compute the zero bias solution to establish the 2-D electron gas in the channel. We also computed a classical zero bias solution for this structure as a reference. The results for the density
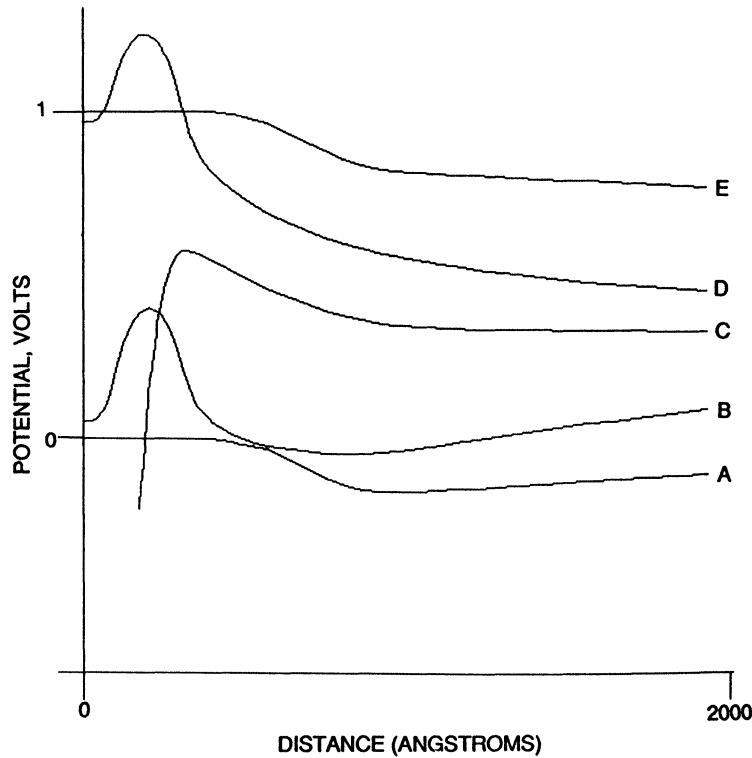
FIGURE 10   Profiles of self consistent potential at selected planes normal to the device surface.

distributions are compared in Figs. 12 and 13. Figure 12 shows a comparison of the classical and quantum corrected density distributions in the direction normal to the device surface, through the channel and extending slightly into the AlInAs buffer layer. The distributions lie in a plane passing through the device half-way between the source contact and the gate recess. Here we note that the classical solution yields higher peak densities and sharp, abrupt interfaces. In comparison, the quantum corrected result shows a continuous density distribution with the peaks reduced in magnitude and a reduction in the depth of the depletion of the planar doped layer. Both results show an initial peak in the InGaAs cap layer at the interface with the gate isolation layer. The second peak occurs in the region of the planar doped layer. Recall that this layer is modeled as a 10 Å wide layer. It is doped to $5 \times 10^{19}/cm^3$ to yield the desired planar density of $5 \times 10^{12}/cm^2$. This layer is heavily depleted. The carriers depleted from the planar doped layer seed the channel and provide the mobile charge for the 2-DEG which is represented by the third peak. Some of the depleted carriers accumulate in the cap layer yielding the peak observed there. With the

exception of the region near the interface between the spacer layer and the channel, the charge distribution across the channel is very similar for both calculations. At the interface between the channel and the buffer layer, we again observe some differences arising from the quantum effects at this abrupt junction. We must also emphasize that these results are on a log scale. The peak densities in the 2-DEG are significantly different. In the quantum corrected result the peak is $7.4 \times 10^{18}/cm^3$. In the classical case the peak density is nearly four times greater at $2.9 \times 10^{19}/cm^3$. However, the total charge in the channel is not altered to as great a degree due to the broader nature of the quantum mechanical distribution. The major significant difference is the level of charge in the spacer layer. Here the classical result shows almost a complete absence of charge whereas the quantum corrected result yields a minimum density in this region of $1.5 \times 10^{18}/cm^3$. This is clearly a significant level of charge and will contribute to conduction in the device. We also note, that in accordance with Eq. (7) the quantum corrected density distribution reflects a continuous variation in E_t while the classical result reflects the discontinuities associated with the barriers at the
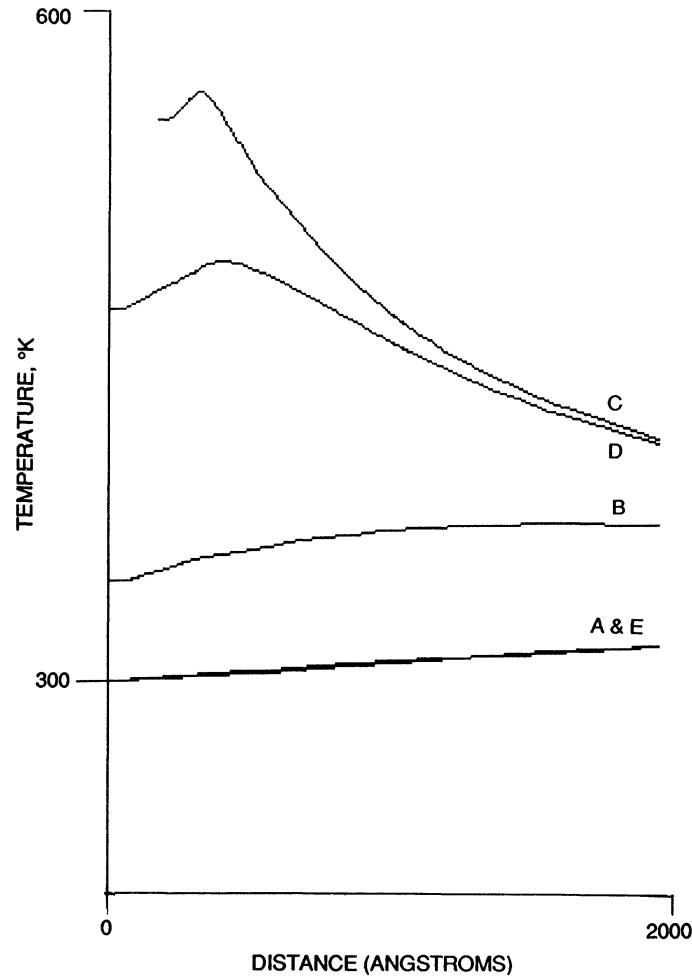
FIGURE 11   Profiles of electron temperature at selected planes normal to the device surface.

material interfaces. The quantum potential is responsible for this as it always tends to smooth out these discontinuities.

Figure 13 shows a similar comparison for a plane dissecting the device normal to the gate contact. We observe the continuous variation in density in the quantum result, increasing from the gate to a peak of $3.2 \times 10^{18}/cm^3$ in the 2-DEG, then decreasing across the channel until reaching the interface with the buffer layer. The classical result exhibits a double peak structure and still clearly reflects the presence of the planar doped layer. The higher peak associated with the 2-DEG is at $1.08 \times 10^{19}/cm^3$. At the gate surface the density is approximately two orders of magnitude greater in the quantum corrected solutions. This is consistent with a lowering of the barrier height by the quantum potential.

The peak density in the 2-DEG is lower in the quantum corrected result than in the classical result. In fact, if we examine a plane parallel to the device

surface passing through the peak density in the 2-DEG from source to drain, the quantum corrected distribution shows a significantly lower peak than the classical result along the entire length of the channel.

Surface plots of the equilibrium density and potential distributions are shown in Fig. 14. Recall that the doping under the contacts has been raised to $1 \times 10^{19}/cm^3$ to give a better representation of the metalization. A wire mesh plot is presented for the potential since it reveals details behind the 2-DEG. It also shows the mesh structure which consists of 107 X points by 136 Z points. We observe in the density plot that the 2-DEG does not extend completely into the contact regions. This is a result of a grading, in the source-drain direction, of the barrier associated with the interfaces between the InGaAs cap layer and channel, and the AlInAs gate isolation layer. This grading is introduced, as previously discussed, to represent the obliteration of these inter-
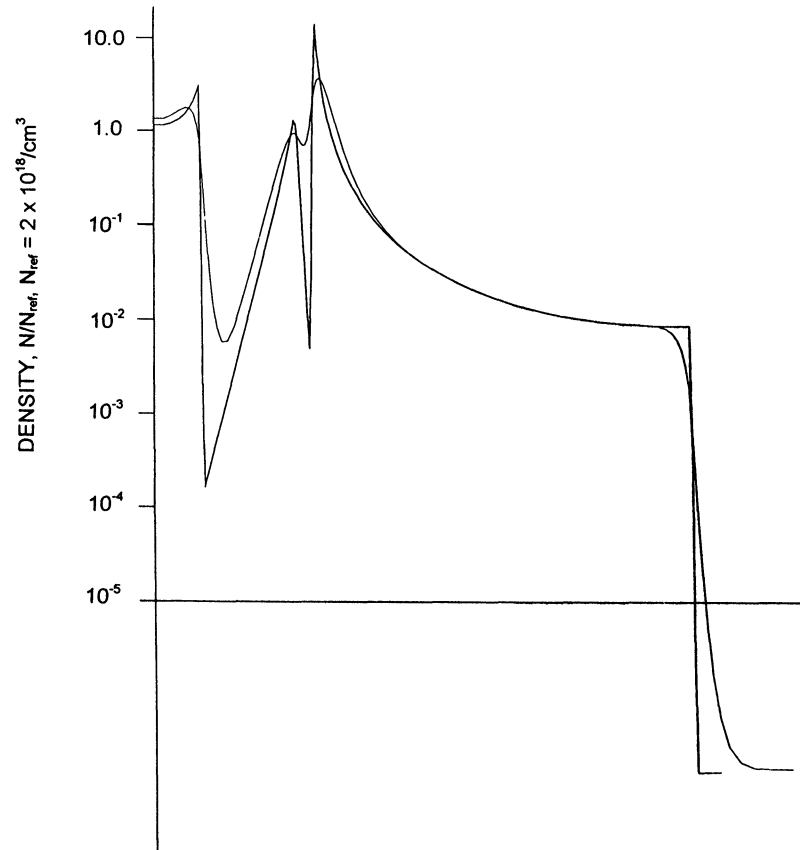
FIGURE 12   Comparison of classical and quantum corrected density distributions in a plane normal to the device surface halfway between the source contact and gate recess.

faces by the metalization. These barriers, and as a result the 2-DEG, are well established before the active gate region of the device is reached. No grading was introduced in the direction normal to the barriers. We also note that as we traverse from the source to the gate there is an increase in the peak density of the 2-DEG in the region of the gate recess. The same effect is also apparent on the drain side of the device. This is a direct result of the gate recess and the absence of the barrier between the cap layer and the isolation layer which is present away from the gate recess. Away from the recess, carriers depleted from the planar doped layer accumulate in both the cap layer and the 2-DEG. The carriers that accumulate in the cap layer are responsible for the first peak shown in the density distribution of Fig. 12. Under the gate recess, carriers from the planar doped layer can accumulate only in the 2-DEG. This has the two-fold effect of increasing the 2-DEG density and reducing the level of depletion of the planar doped layer in the region. Directly under the gate contact the peak density is reduced due to depletion by the Schottky barrier.

The potential surface reveals a large rise in potential from the source extending to the drain where it returns to the same level as at the source. This spike in potential is associated with the depletion of the planar doped layer and establishment of the 2-DEG in the channel. Under the gate recess the slight increase in potential is a result of the higher density of the 2-DEG discussed previously. The presence of the gate contact is apparent in the lower potential in the depletion region.

Figure 15 shows an enlarged wire mesh plot of the quantum potential for this case. The region shown in the figure extends from an area just to the source side of the gate recess to the center of the gate contact. The region extends only slightly into the channel. At the left end of the figure we see a one-dimensional distribution of the quantum potential exists from the device surface into the channel. The quantum potential is zero at the device surface and decreases to a local minimum at the interface between the cap layer and the gate isolation layer. Across the interface, in the AlInAs, a maximum is reached. This quickly decays to zero, then after a
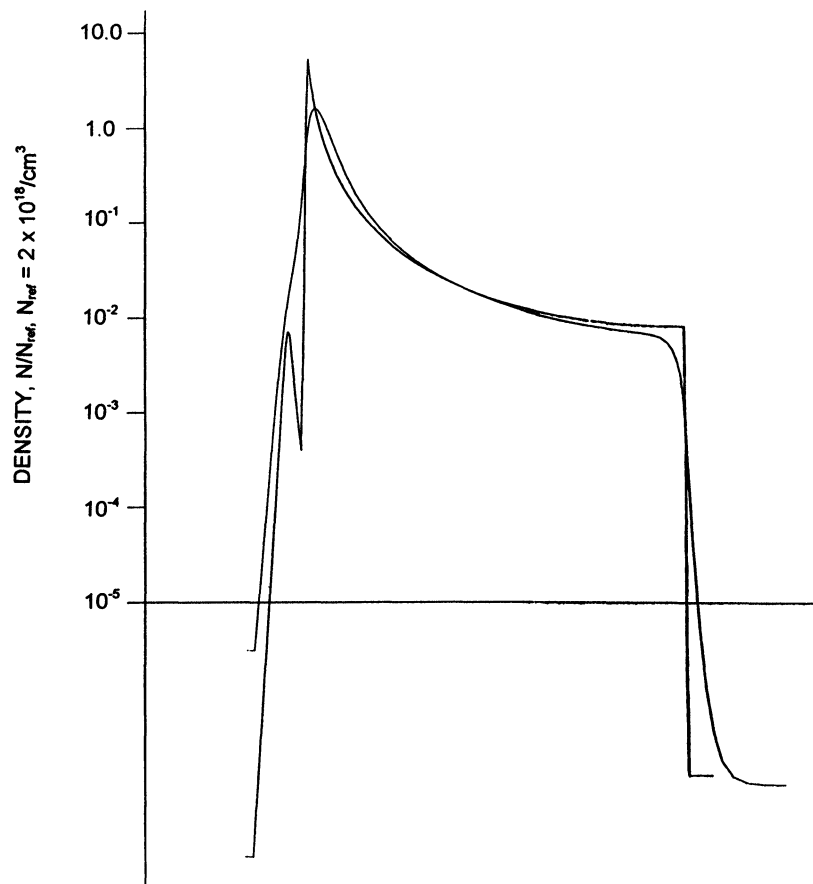
FIGURE 13   Comparison of classical and quantum corrected density distributions in a plane normal to the device surface under the gate contact.
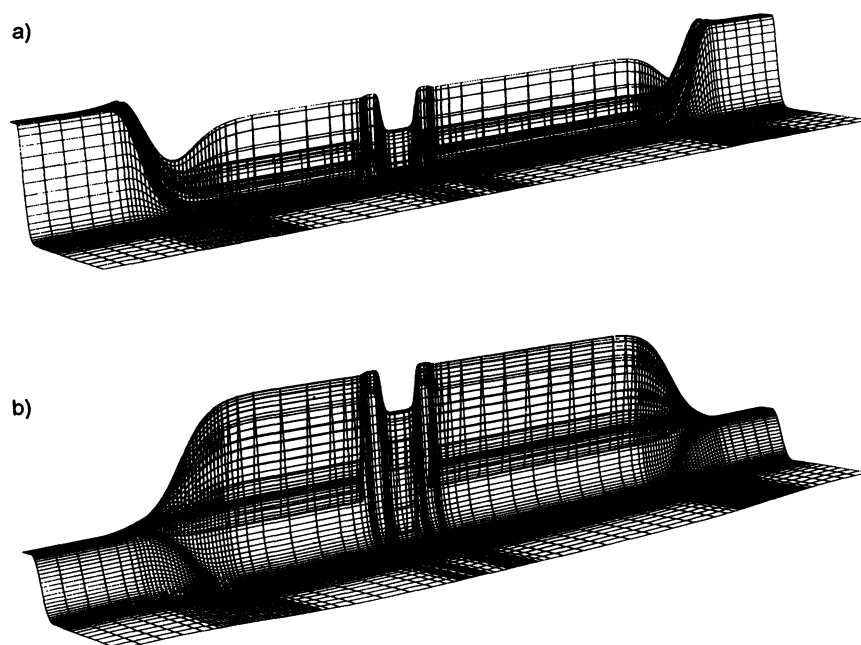

FIGURE 14   Surface plots of a) electron density and b) potential for equilibrium solution.
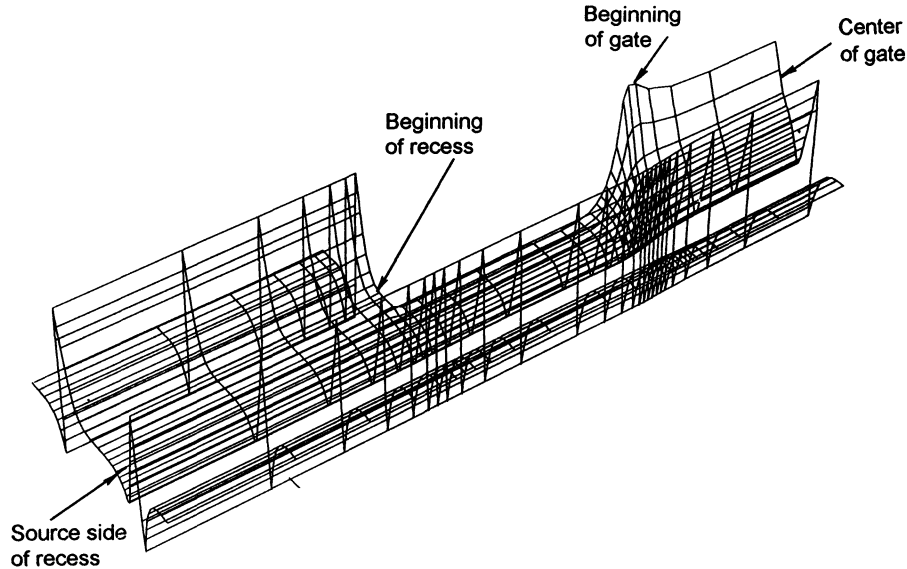
FIGURE 15   Blowup of surface plot of equilibrium quantum potential in region between the center of the gate contact and to the source side of the gate recess.

dip in the region of the planar doped layer, the quantum potential rises to a second maximum on the AlInAs side of the interface with the channel. Crossing the barrier into the channel we observe a second minimum on the InGaAs side of the barrier which quickly decays to zero. We also note that the length scales over which the quantum potential varies are consistent with our analysis and are on the order of tens of angstroms.

In the region of the gate recess, we see that the quantum potential still behaves one-dimensionally. The recess itself does not alter the distribution significantly. However, as we approach the gate contact from the side, we observe a rapid rise in the quantum potential. The same effect is noted as we approach the gate in the direction normal to the gate surface. This is because the Schottky contact represents another barrier which the quantum mechanical effects try to lower. The effects are highly localized and have little influence on the distribution of the quantum potential at the channel interface. A slight increase in the quantum potential in the region of the channel interface is noted as we move in the source to drain direction, but this is not significant compared to the variations normal to the surface.

When a bias is applied to the structure the distribution of density, potential, velocity along the channel, and temperature appear as shown in Fig. 16. This result is for the case of $V_{ds} = 0.5$, and $V_{gs} = 0.4$ volts. The most pronounced effects observed in the

density surface are the increase in the density at the source side of the gate recess, the distortion of the depletion region with lower density at the drain end of the gate and the reduction of the density peak on the drain side of the gate recess. These three effects are all the result of the gate recess and the applied bias. As a result of the constriction caused by the gate recess, additional electrons accumulate on the source side of the gate in this region. They accumulate there because they are about to enter the high resistance region under the gate. Directly under the gate the electrons accelerate, further reducing the density in the 2-DEG. This effect is typical of the depletion under a Schottky contact. To the drain side of the gate, the peak in the 2-DEG caused by the gate recess is eliminated because the electrons in this region are quickly drawn away along the highly conductive path to the drain.

The potential distribution reflects these effects and shows that relatively uniform fields exist along the 2-DEG. The field in the cap layer, on both sides of the gate recess, is nearly zero. The small fields which do exist in the cap layer in the source-drain direction balance a small accumulation of charge on the source side of the gate recess and depletion on the drain side. This result is similar to that of our preliminary device structure. We reiterate that the sloshing charge in the cap layer will affect the parasitic capacitance of such devices. Under the gate we observe that the potential rises to a higher value than in our preliminary structure. This is a direct
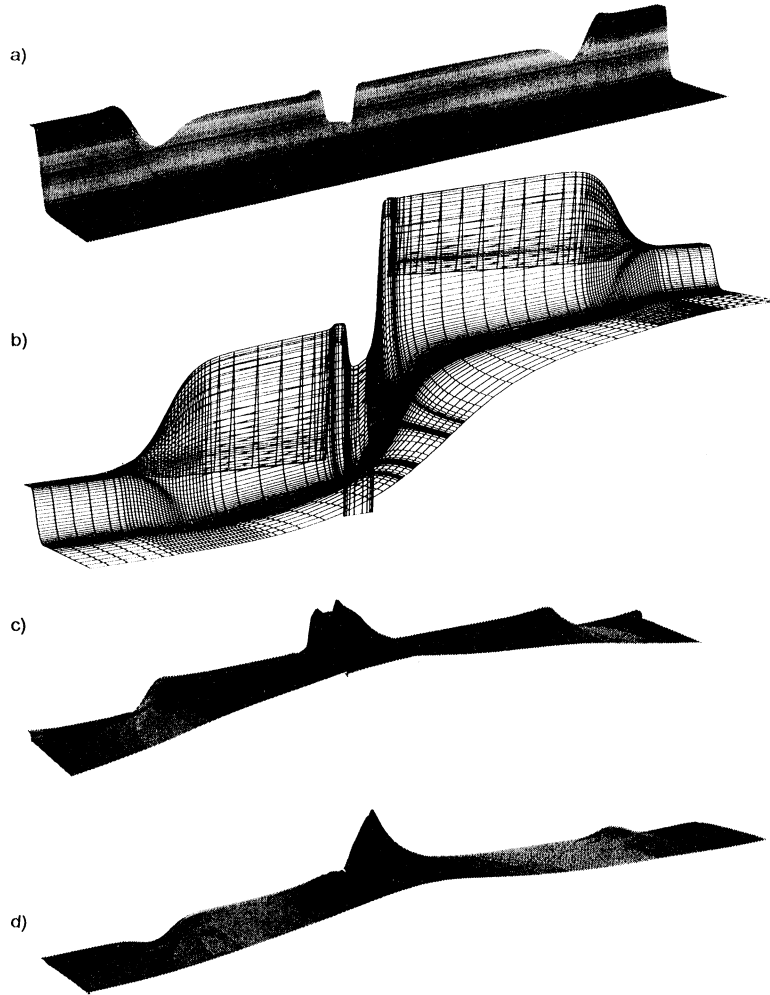
FIGURE 16   Surface plots of a) density, b) potential, c) velocity and d) temperature at $V_{ds} = 0.5$, $V_{gs} = 0.4$ volts.

result of the increased density of the 2-DEG when the planar layer is correctly modeled. The total density of dopants in the isolation layer is nearly the same in both cases, $4 \times 10^{12}/\text{cm}^2$ in the preliminary structure and $5 \times 10^{12}/\text{cm}^2$ in the real structure. However, due to the proximity of the planar doped layer to the channel, most of the charge seeds the 2-DEG whereas in the preliminary structure most of the charge accumulated in the cap layer. Additionally, since the isolation layer is only 100 Å thick under the gate recess, only half as much charge was available to seed the 2-DEG in the preliminary structure. Thus, as is evident from the density surface plot, the channel is much more highly conductive when the planar layer is correctly simulated.

Figure 16 also shows surface plots of the velocity along the channel and the electron temperature. The velocity surface shows the acceleration of electrons as they enter the 2-DEG. Between the source

and gate, and the gate and drain the electron velocity is almost constant in the 2-DEG. As we move across the channel into the buffer layer, the velocity shows a smooth increase and then a smooth decrease moving from source to drain. Directly under the gate we observe a rapid acceleration in response to the lower density arising from the gate depletion layer. A peak in the velocity occurs at the drain end of the gate contact. The peak approaches 1.35×10^7 cm/sec. While difficult to see in this figure, the velocity in the cap layer is zero. An indication of this is apparent at the source end of the figure.

Finally, we examine the electron temperature surface. Here we note a minor cooling of electrons as they enter the 2-DEG. This results because of an exchange of thermal energy to kinetic energy in this region. The thermal response lags the acceleration spatially. The effect is short lived, however, and the electrons then rise slightly in temperature in the
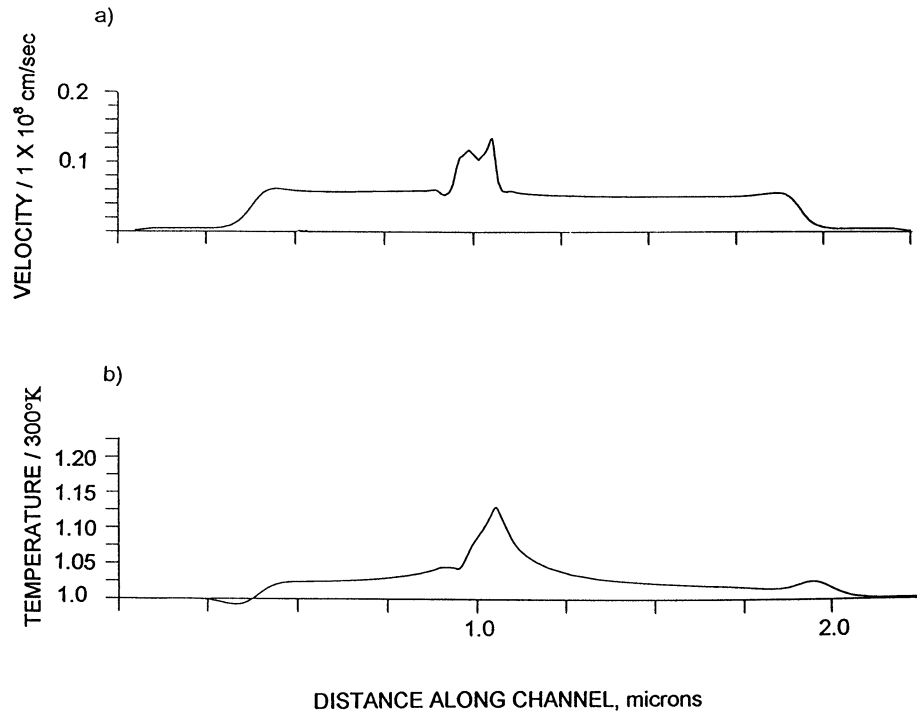
FIGURE 17   Velocity, a) and temperature, b) along the peak in the 2-DEG and $V_{ds} = 0.5$, $V_{gs} = 0.4$.
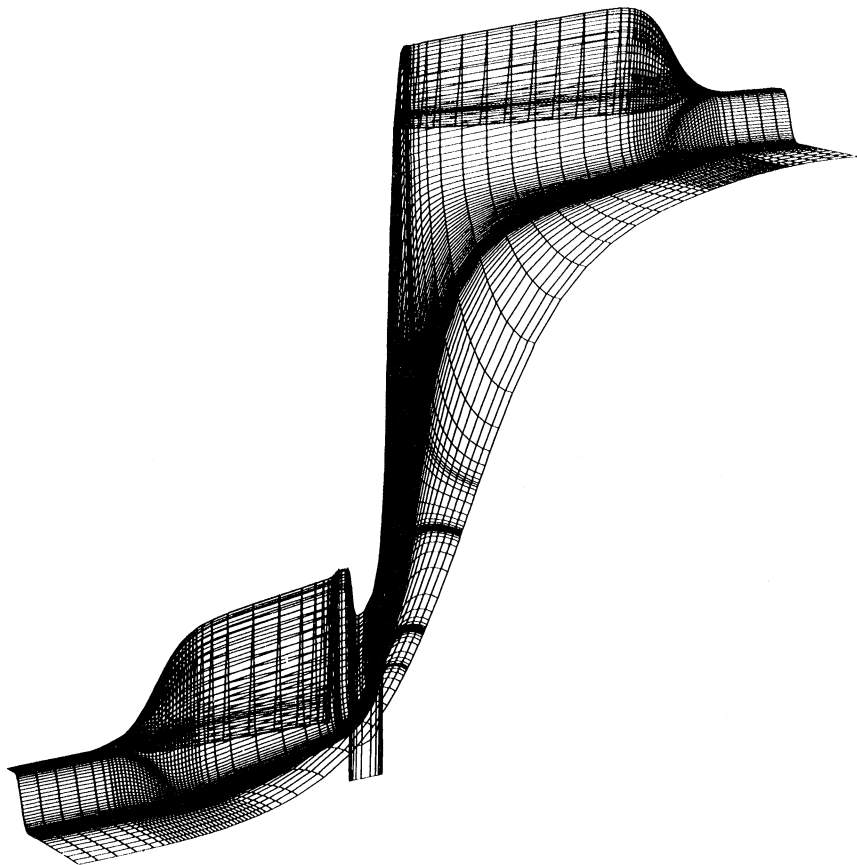

FIGURE 18   Surface plot of potential at a bias of $V_{ds} = 1.5$, $V_{gs} = 0.4$ volts.

established region of the 2-DEG. The electron temperature rises considerably under the gate. Upon reaching the drain end of the gate they then decelerate, cool and maintain a constant temperature until they reach the drain. The maximum temperature under the gate reaches 340°K at this bias level. Upon further deceleration, as the electrons exit the 2-DEG and arrive at the drain, an inverse of the effect observed at the source is shown. Here the excess electron kinetic energy is initially converted to thermal energy as the deceleration begins and the temperature rises slightly. As the electrons continue to slow, the temperature relaxes to the lattice temperature. The peak temperature is lower than in the preliminary structure because of the greater conductivity of the channel.

The electron velocity and temperature variations in the plane of the peak density in the 2-DEG are shown in Fig. 17 at the same bias condition. Here we can clearly observe the acceleration of electrons from the source region into the 2-DEG where the velocity is near 0.6×10^7 cm/sec. The slight dip in the velocity is in the region where the density increased due to the beginning of the gate recess. The electrons then accelerate sharply as they enter the gate depletion region. A final rapid acceleration to 1.35×10^7 cm/sec occurs in the region of minimum density and maximum electric field. The electrons then decelerate to a level somewhat below 0.6×10^7 cm/sec. A slight acceleration occurs just before the electrons decelerate at the drain metalization.

The temperature distribution shows clearly the initial drop below 300°K as the electrons enter the 2-DEG. Once in the 2-DEG the temperature rises to 310°K. The electron temperature continues to

rise to about 315°K before encountering the region under the gate. Here the temperature peaks at 340°K. The electrons then cool as they approach the drain. The minimum in temperature and slight rise before cooling in the drain region coincide with the small velocity peak and initiation of deceleration just before the drain is reached.

When the drain bias is increased the distributions of potential, density, velocity, etc. are qualitatively similar. Higher velocities and temperatures occur under the gate, and there is deeper depletion at the drain edge of the gate. As an example of such results we present the potential distribution at $V_{ds} = 1.5$ volts and $V_{gs} = 0.4$ volts in Fig. 18. We note that the distribution of potential on the source and drain sides of the gate is almost identical to that at $V_{ds} = 0.5$ volts due to the high conductivity of the 2-DEG. Most of the increase in potential is applied to overcome the high resistance region at the drain side of the gate. The velocity reaches a peak value of 1.41×10^7 cm/sec and the temperature reaches 365°K in this region.

The previous discussion and figures provide a vivid description of the internal dynamics of the present HEMT structure. In Fig. 19 we show the I-V characteristics for this device at gate bias levels of 0.0, 0.2 and 0.4 volts. The results are compared to experimental measurements for the same structure. Here we observe that there is good qualitative agreement between prediction and measurement. The major discrepancy is in the device transconductance and is reflected by the higher current levels at lower gate bias. The predictions yield a transconductance of 711 ms/mm at $V_{gs} = 0.2$ and $V_{ds} = 1.5$ volts with a corresponding value of the unity gain
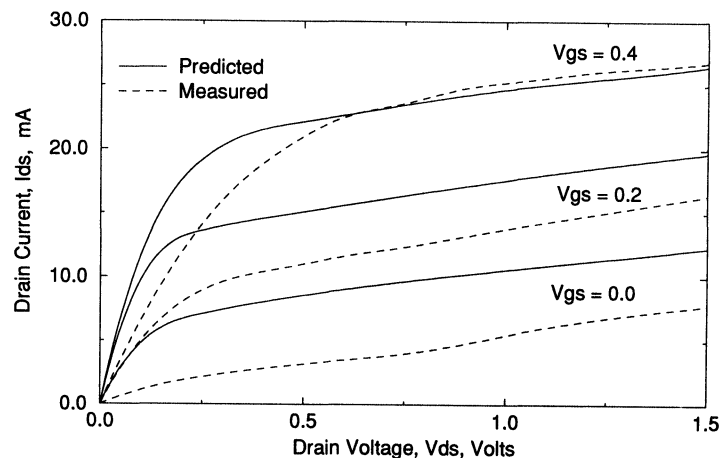

FIGURE 19   Comparison between predicted and measured I-V characteristics.

cutoff frequency of $f_t = 134$ GHz. The experimental measurements yield $g_m = 1075$ ms/mm with $f_t = 196$ GHz. Both are about 50% higher than predicted. There are several potential sources of these discrepancies. Two of these are the density of the planar doped layer and the saturated drift velocity used in determining the relaxation time. The transconductance can be increased by increasing either the saturated drift velocity or increasing the density in the planar doped layer. Our work has shown that altering V_{sat} has little effect on the device capacitance, thus as the transconductance increases, so does f_t. However, the current levels will also increase. On the other hand, changing the density of the planar doped layer affects both the transconductance and the device capacitance. As the planar layer doping is decreased, both the transconductance and device capacitance are reduced, as is the drain current. The dependence of the relationship between device capacitance and transconductance is, however, highly nonlinear as there are other sources of capacitance within the device. A third factor is the depth of the gate recess. In the actual device there is a degree of uncertainty associated with the precise depth of the gate recess. It is known that if the gate recess is deeper there is better control and the device transconductance increases. Thus, we see that there are a number of parameters associated with the device which will influence performance and may account for the observed discrepancies between the predicted and measured characteristics. As long as uncertainties exist with regard to parameters (such as the planar layer doping level and gate recess depth), it will be difficult to access the limitations of our predictive capabilities. However, such predictive capability does provide a means of investigating the range of the performance variations associated with these uncertainties. We are presently examining these factors and their impact on device performance.

CONCLUSIONS

A series of one- and two-dimensional simulations have been performed using the QHD equations. The conclusions drawn from the one-dimensional results are that under the equilibrium conditions considered here the results obtained from the QHD equations are in good agreement with those obtained from the Liouville equation. We have also investigated the effect of the multiplicative factor of the Bohm potential appearing in the governing equations. As this factor becomes smaller (i.e., 1/9) the peak density on the narrow gap side of a single heterojunction increased towards the classical limit. Best agreement with the Liouville equation is obtained for a value of 1/3. The results also indicated that in the classical limit the charge sheet density in the channel of an HEMT would be over predicted.

The results for a 2-D quantum well HEMT show the importance in including quantum mechanical effects in such simulations. The results indicate that for structures such as this HEMT the quantum potential plays a significant role in the creation of the 2-DEG in the channel. As with the one-dimensional results, the two-dimensional results indicate that the charge sheet density in the 2-DEG is over-predicted without quantum corrections. The effects are primarily limited to one-dimensional effects normal to the heterojunctions, but in the region of the gate surface the quantum potential will have a profound two-dimensional effect and will contribute to tunneling of electrons through the gate contact. The quantum mechanical contribution must be included in the simulation of such structures if reliable results are to be expected.

References

[1] D. Bohm, "Suggested New Interpretation of the Quantum Theory in Terms of Hidden Variables, I," *Phys. Rev.*, vol. 85, p. 166, 1952.

[2] E. Wigner, "On the Quantum Correction for Thermodynamic Equilibrium," *Phys. Rev.*, vol. 40, pp. 749–759, 1932.

[3] H.L. Grubin and J.P. Kreskovsky, "Quantum Moment Balance Equations and Resonant Tunnelling Structures," *Solid State Electronics*, p. vol. 32, no. 12, pp. 1071, 1075, 1989.

[4] C.L. Garner, "The Classical and Quantum Hydrodynamic Models," *Proceedings of the International Workshop on Computational Electronics*, pp. 25–36, 1993.

[5] H.L. Grubin, T.R. Govindan, and J.P. Kreskovsky, "Transport Via the Liouville Equation and Moments of Quantum Distribution Functions," *Solid State Electronics*, vol. 36, no. 12, pp. 1697–1709, 1993.

[6] W.R. Briley and H. McDonald, "On the Structure and Use of Linearized Block Implicit Schemes," *J. Comp. Phys.*, vol. 34, no. 1, pp. 54–73, 1980.

[7] J.P. Kreskovsky and H.L. Grubin, "Application of LBI Techniques to the Solution of the Transient, Multidimensional Semiconductor Equations," *J. Comp. Phys.*, vol. 68, no. 2, pp. 420–461, 1987.

[8] J.P. Kreskovsky, M. Meyyappan and H.L. Grubin, "The Moments of the Boltzmann Transport Equation as Applied

to the Gallium Arsenide Permeable Base Transistor," *Fundamental Research on the Numerical Modeling of Semiconductor Devices and Processes*, pp. 99–106, Boole Press, 1987.

[9] J. Douglas and J.E. Gunn, "A General Formulation of Alternating Direction Methods, Part I, Parabolic and Hyperbolic Problems," *Numer. Meth.*, vol. 6, pp. 428–453, 1964.

[10] J-R. Zhou and D.K. Ferry, "Simulation of Ultra-Small GaAs MESFET Using Quantum Moment Equations," *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 473–478, 1992.

[11] J.R. Zhou and D.K. Ferry, "Simulation of Ultra-Small GaAs MESFET Using Quantum Moment Equations II: Velocity Overshoot," *IEEE Trans. Electron Devices*, vol. 39, no. 8, pp. 1793–1796, 1992.

[12] Y-C. Pao, C.K. Nishimoto, R. Majidi-Ahy, J. Archer, N.G. Bechtel and J.S. Harris, Jr., "Characterization of Surface-Doped $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As/InP$ High Electron Mobility Transistors," *IEEE Trans. Electronic Devices*, vol. 37, no. 10, pp. 2165–2170, 1990.

Biographies

JOHN P. KRESKOVSKY received the B.S. degree in 1969 and the M.S. degree in 1973 in the field of mechanical engineering from Pennsylvania State University. He joined the staff of United Technologies Research Center in 1973 where he was responsible for the development of computational producers for solving complex fluid dynamics problems using the Navier-Stokes equations. In 1977 he joined Scientific Research Associates as a co-founder. Beginning in 1981, Mr. Kreskovsky's efforts have been focused on developing and applying simulation techniques to problems in solid state device physics. His current interests include device research and simulation at both the classical and quantum levels with emphasis on compound semiconductors.

HAROLD L. GRUBIN was born in Brooklyn, New York in 1939. He received the Ph.D. from the Polytechnic Institute of Brooklyn in 1967. In 1966, Dr. Grubin joined United Technologies Research Center and began his studies of hot electron transport effects in semiconductor devices. In 1981, he joined Scientific Research Associates, Inc. His research has included analytical and numerical simulation methods as applied to the physics and operation of transferred electron compound and elemental semiconductor devices. His work includes hot carrier effects, electron and hole transport, two terminal devices, MESFETs, MOSFETs, JFETs, bipolar transistors, HBTs, HEMTs, large and small signal time dependent behavior, equivalent circuit modeling, one-, two- and three-dimensional numerical simulations, rectangular and non-rectangular shaped devices, radiation effects, simulations of Czochralski growth of silicon and gallium arsenide single crystals, Boltzmann transport and quantum transport in tunneling and superlattice structures. Dr. Grubin is the Co-author of: *The Gunn-Hilsum Effect*, Academic Press and *The Physics of Instabilities in Solid State Electron Devices*, Plenum Press (1992). Dr. Grubin was Co-editor of the IEEE Transactions on Electron Devices, *Special Issue on Short Channel Effects in Devices* (1981); *The Physics of Submicron Structures*, Plenum Press (1984), and *The Physics of Submicron Semiconductor Devices*, Plenum Press (1988). He is International Co-editor of the Wiley series Design and Measurement in Electronic Engineering. Dr. Grubin is an Adjunct Professor in the Department of Electrical and Computer Engineering at North Carolina State University.